

У статті описано основну проблему некоректної обробки текстових даних інформаційними системами, виявлено недоліки існуючих способів представлення знань. Запропоновано узагальнену форму логіко-лінгвістичної моделі як універсального засобу вилучення знань з текстової інформації. Виведено типові форми узагальненої логіко-лінгвістичної моделі для конкретних видів речень природної мови

Ключові слова: природна мова, логіко-лінгвістична модель, текстова інформація, обробка тексту, формалізація

В статье описана основная проблема некорректной обработки текстовых данных информационными системами, выявлены недостатки существующих способов представления знаний. Предложена обобщенная форма логико-лингвистической модели как универсальное средство извлечения знаний из текстовой информации. Выведены типичные формы обобщенной логико-лингвистической модели для конкретных видов предложений естественного языка

Ключевые слова: естественный язык, логико-лингвистическая модель, текстовая информация, обработка текста, формализация

УДК 004.519.7 (045)

DOI: 10.15587/1729-4061.2015.43049

ФОРМУВАННЯ ЛОГІКО- ЛІНГВІСТИЧНИХ МОДЕЛЕЙ ТИПОВИХ ПРИРОДНОМОВНИХ КОНСТРУКЦІЙ

А. І. Вавіленкова

Кандидат технічних наук, доцент
Кафедра комп'ютеризованих систем управління
інституту комп'ютерних
інформаційних технологій
Національний авіаційний університет
пр. Космонавта Комарова, 1,
м. Київ, Україна, 03068
E-mail: a_vavilenkova@mail.ru

1. Вступ

В основній своїй більшості дані у інформаційних системах представлені у вигляді текстової інформації. Текст являється однією з основних форм обміну інформацією у суспільстві. Текстова інформація у різних формах складає значну долю ресурсів інформаційних систем. Тому створення та розвиток інформаційних технологій обробки текстової інформації викликає інтерес на всіх етапах розвитку інформаційних систем.

Обробка текстової інформації лежить в основі сучасного програмного забезпечення, яке розпізнає та інтерпретує природну мову. Уже отримано задовільні результати в здійсненні морфологічного, синтаксичного та частково семантичного аналізу [1], проте складною задачею залишається розуміння змісту текстової інформації [2]. Сучасні засоби автоматичної обробки природномовних даних не задовольняють потреби користувачів, тому на перший план виходить інтелектуалізація засобів пошуку та створення релевантних моделей обробки інформації. Тому сьогодні актуальною є задача розробки формального апарату, що дозволить би вилучати з текстової інформації знання.

Одним із підходів до вирішення основної проблеми автоматичної обробки текстової інформації є підхід, заснований на формуванні змістовних моделей текстових документів на базі використання математичного апарату логіки предикатів.

2. Аналіз літературних даних і постановка проблеми

Обробка природної мови лежить в основі сучасного програмного забезпечення, яке розпізнає та ін-

терпретує природну мову. Уже отримано задовільні результати в здійсненні морфологічного, синтаксичного та частково семантичного аналізу, проте складною задачею залишається розуміння змісту текстової інформації [3].

Основною проблемою на шляху коректної обробки текстових даних інформаційними технологіями є відсутність єдиного підходу до аналізу текстової інформації [4, 5]. Системи інтелектуальної обробки текстової інформації використовують результати існуючих програмних засобів попередньої обробки текстів, внаслідок чого похибка функціонування створеної аналітичної системи збільшується за рахунок похибки існуючого програмного забезпечення. Використання різних формальних моделей представлення текстової інформації на різних етапах обробки приводить до багаторазового перетворення текстової інформації на природній мові у моделі представлення знань і навпаки.

Проблемі представлення знань присвячено багато літературних джерел. Починаючи з витоків проблеми формалізації природної мови (праці Поспелова Д. В [6], Осуга С., Саєкі Ю. [7]) і до теперішнього етапу обробки текстової інформації (Джарратано Д. [8], Широков В. А. [9], Ікеда М. [10]), вченим вдалося винайти п'ять формальних моделей представлення знань: продукційні, фреймові, логічні, семантичні мережі та онтології. Проте жодна з цих моделей не може в повній мірі відобразити зміст тексту на природній мові.

Методи інтелектуального аналізу текстової інформації Data Mining допомагають вирішити багато задач, з якими стикається аналітик, зокрема, розроблені та удосконалені методи класифікації та кластеризації описані багатьма вченими [11, 12].

Інформаційні технології інтелектуального аналізу позиціонуються як механізми, що використовують для роботи сукупність законів, методів та засобів отримання, зберігання, передачі, поширення, перетворення інформації про мову та закони її функціонування за допомогою комп'ютерів. Сьогодні комп'ютерні лінгвісти [13] у більшості випадків використовують для обробки текстів статистичні методи без аналізу глибинних структур, тоді як ринок інформаційних технологій потребує нового програмного продукту, який базувався б на змістовному аналізі моделей представлення знань.

Тому для сфери інтелектуальної обробки даних актуальною є задача розробки формального апарату, що дозволив би вилучати з текстової інформації зміст. У свою чергу, ринок інформаційних технологій потребує створення нового програмного продукту, який базувався б на змістовному аналізі моделей представлення знань.

3. Мета та задачі дослідження

Метою досліджень є створення узагальненої логічної моделі представлення знань та її типових форм за використанням апарату логіки предикатів.

Для досягнення поставленої мети необхідно вирішити такі завдання:

- проаналізувати концептуальні відношення у реченнях природної мови;
- виявити можливі типи логічних зв'язків у синтаксичних конструкціях речень природної мови;
- сформулювати узагальнену логіко-лінгвістичну модель для речень природної мови довільної складності;
- вивести типові форми узагальненої логіко-лінгвістичної моделі.

4. Узагальнена логіко-лінгвістична модель представлення речень природної мови довільної складності

Так як речення – це слово або група слів, за допомогою яких відображається думка, і воно характеризується смисловою завершеністю, граматичною організацією та предикативністю, то кожне речення природної мови можна інтерпретувати як логічну формулу.

Фактично логіко-лінгвістична модель є зв'язкою між елементами формальної логіки та синтаксичною структурою речень природної мови. Просте речення у формальній логіці – це атомарний предикат, складне речення – складне логічне висловлювання, сукупність атомарних предикатів поєднаних логічними зв'язками [14].

Виходячи із цього, сформовано уніфіковану форму логіко-лінгвістичної моделі речення природної мови:

$$L^S = \bigwedge_{p \in P^S} \bigwedge_{h \in H_p^S} L_p^S(h), \quad (1)$$

$$L_p^S(h) = \bigwedge_{x \in X_p^S(h)} \bigwedge_{g \in G_p^S(x,h)} L_p^S(x,g,h), \quad (2)$$

$$L_p^S(x,g,h) = \bigwedge_{y \in Y_p^S(x,g,h)} \bigwedge_{q \in Q_p^S(x,g,y,h)} L_p^S(x,g,y,q,h), \quad (3)$$

$$L_p^S(x,g,y,q,h) = \bigwedge_{z \in Z_p^S(x,g,y,q,h)} \bigwedge_{r \in R_p^S(x,g,y,q,z,h)} L_p^S(x,g,y,q,z,r,h), \quad (4)$$

де S – речення природної мови; p – відношення, що пов'язує суб'єкти, об'єкти та предмети відношень у реченні S ; $p \in P^S$ – множина відношень, що входять до речення S ; h – характеристика p -го відношення речення S ; $h \in H_p^S$ – множина характеристик p -го відношення у реченні S ; $L_p^S(h)$ – предикат (предикативний вираз) [8], який описує p -е відношення з h -ю характеристикою і пов'язує суб'єкти, об'єкти та предмети відношення p в реченні S ; x – суб'єкт речення S , $x \in X_p^S(h)$ – множина суб'єктів, що пов'язані з об'єктами речення S p -им відношенням, яке володіє h -ю характеристикою; g – характеристика суб'єкта x речення S , $g \in G_p^S(x,h)$ – множина характеристик суб'єкта $x \in X_p^S(h)$; $L_p^S(x,g,h)$ – предикат (предикативний вираз), який описує p -е відношення з h -ю характеристикою між суб'єктом $x \in X_p^S(h)$ з характеристикою $g \in G_p^S(x,h)$, об'єктами та предметами p -го відношення в реченні S ; y – об'єкт речення S , $y \in Y_p^S(x,g,h)$ – множина об'єктів, що пов'язані з суб'єктами речення S p -им відношенням, яке володіє h -ю характеристикою; q – характеристика об'єкта y речення S , $q \in Q_p^S(x,g,y,h)$ – множина характеристик об'єкта $y \in Y_p^S(x,g,h)$; $L_p^S(x,g,y,q,h)$ – предикат (предикативний вираз), який описує p -е відношення з h -ю характеристикою між суб'єктом $x \in X_p^S(h)$ з характеристикою $g \in G_p^S(x,h)$ і об'єктом $y \in Y_p^S(x,g,h)$ з характеристикою $q \in Q_p^S(x,g,y,h)$ та предмети p -го відношення в реченні S ; z – предмет p -го відношення речення S , $z \in Z_p^S(x,g,y,q,h)$ – множина предметів p -го відношення, яке володіє h -ю характеристикою, між суб'єктом $x \in X_p^S(h)$ з характеристикою $g \in G_p^S(x,h)$ та об'єктом $y \in Y_p^S(x,g,h)$ з характеристикою $q \in Q_p^S(x,g,y,h)$; r – характеристика предмета p -го відношення речення S , $r \in R_p^S(x,g,y,q,z,h)$ – множина характеристик предмета $z \in Z_p^S(x,g,y,q,h)$; $L_p^S(x,g,y,q,z,r,h)$ – простий, неділимий предикат, який описує частину речення, яка має закінчений зміст, та відображає в реченні S p -е відношення з h -ю характеристикою між суб'єктом $x \in X_p^S(h)$ з характеристикою $g \in G_p^S(x,h)$ і об'єктом $y \in Y_p^S(x,g,h)$ з характеристикою $q \in Q_p^S(x,g,y,h)$, предмет якого $z \in Z_p^S(x,g,y,q,h)$ володіє характеристикою $r \in R_p^S(x,g,y,q,z,h)$.

Кількість v^S частин речення S , що мають закінчений зміст і описуються простим предикатом $L_p^S(x,g,y,q,z,r,h)$, розраховується за формулами:

$$v^S = \sum_{p \in P^S} \sum_{h \in H_p^S} v_p^S(h),$$

$$v_p^S(h) = \sum_{x \in X_p^S(h)} \sum_{g \in G_p^S(x,h)} v_p^S(x,g,h),$$

$$v_p^S(x,g,h) = \sum_{y \in Y_p^S(x,g,h)} \sum_{q \in Q_p^S(x,g,y,h)} v_p^S(x,g,y,q,h),$$

$$v_p^S(x,g,y,q,h) = \sum_{z \in Z_p^S(x,g,y,q,h)} |R_p^S(x,g,y,q,z,h)|.$$

Тобто кількість простих предикатів у формулі (1) дорівнює сумі можливих комбінацій узгодження суб'єктів, об'єктів, предметів, відношень між ними та їх характеристик.

Логіко-лінгвістична модель L^S речення S відображається сукупністю формул (1)–(4) та формально представляє собою послідовність восьми кон'юнкцій, що входять до цих формул. Тобто логіко-лінгвістична модель L^S є множиною простих предикатів $L_p^S(x, g, y, q, z, r, h)$, кількість яких дорівнює v^S .

Складовими частинами речення є *словосполучення* – синтаксичні одиниці, що утворюються поєднанням двох або більше повнозначних слів, одне з яких головне, а друге – залежне. Граматичний зв'язок слів у словосполученні здійснюється формами слів за допомогою закінчень та прийменниково-відмінкових форм. В залежності від цього розрізняють ти типи концептуальних відношень між словами:

– *узгодження* – логічний зв'язок, при якому залежне слово стоїть у тій же формі, в якій стоїть головне;

– *керування* – логічний зв'язок слів, при якому залежне слово стає в тому відмінку, якого вимагає головне слово, при зміні головного слова залежне залишається водній і тій же формі;

– *прилягання* – логічний зв'язок, при якому залежне слово приєднується до головного тільки за змістом.

Для виявлення концептуальних відношень між словами в реченні природної мови використовується система продукцій, що представляє собою набір правил типу «modus ponens». Для української мови таких правил нараховується тридцять два [2].

Кожне речення містить концептуальні зв'язки, що інтерпретуються за допомогою набору словосполучень. Таким чином, простий предикат $L_p^S(x, g, y, q, z, r, h) = p(x_1, g_1, y_1, q_1, z_1, r_1, h_1)$ сформовано на основі функціональних зв'язків між концептами (компонентами логіко-лінгвістичної моделі):

– *додаткові* – між присудком і додатком: p та y (відношення керування); між додатком і додатком: y та z (відношення керування);

– *обставинні* – між присудком і обставиною: p та h (відношення прилягання);

– *означальні* – між означенням і підметом: x та g (відношення узгодження); між означенням і додатком: y та q (відношення узгодження), z та r (відношення узгодження).

Логічні зв'язки у реченнях природної мови визначаються за допомогою сполучних слів та знаків пунктуації [2]. Якщо речення описує одночасність, приєднання, сумісність дій або станів, а елементи хоча б однієї з множин логіко-лінгвістичної моделі мають однакові граматичні параметри, то в логіко-лінгвістичній моделі буде використано логічну операцію кон'юнкції $\&$. Якщо концепти речення протиставляються, зіставляються, взаємно виключаються або чергуються, то в логіко-лінгвістичній моделі буде використано логічну операцію диз'юнкції \vee . Якщо одне просте речення уточнює друге, то між ними в логіко-лінгвістичній моделі буде застосована логічна операція імплікації \rightarrow .

5. Результати побудови типових випадків логіко-лінгвістичних моделей

Залежно від типу речення природної мови, уніфікована форма логіко-лінгвістичної моделі (1)–(4) може приймати такі форми.

1) Просте речення природної мови містить лише одну частину, що має закінчений зміст, тобто

$$\begin{aligned} |P^S| = 1; |H^S| = 1; |X_p^S(h)| = 1; |G_p^S(x, h)| = 1; |Y_p^S(x, g, h)| = 1; \\ |Q_p^S(x, g, y, h)| = 1; |Z_p^S(x, g, y, q, h)| = 1; |R_p^S(x, g, y, q, z, h)| = 1. \end{aligned}$$

Внаслідок цього логіко-лінгвістична модель (1)–(4) зводиться до простого предикату $L^S = L_p^S(x, g, y, q, z, r, h)$, а кількість частин речення S , що мають закінчений зміст $v^S = |R_p^S(x, g, y, q, z, h)| = 1$.

Наприклад, для простого речення природної мови «Цікава лекція заохочує студентів до подальшого навчання» логіко-лінгвістична модель буде мати вигляд:

$$L(S) = L^S = p_1(x_1, g_1, y_1, q_1, z_1, r_1, h_1) = p_1(x_1, g_1, y_1, 0, z_1, r_1, 0),$$

$$L(S) = L^S = \text{заохочує (лекція, цікава, студентів, 0, навчання, подальшого, 0)}.$$

2) Логіко-лінгвістична модель простого речення, ускладненого однорідними членами, приймає різні форми залежно від синтаксичних ролей, які виконують у ньому однорідні члени.

А) Якщо множина відношень P^S , що пов'язують суб'єкти, об'єкти та предмети відношень у реченні S містить декілька елементів, то

$$\begin{aligned} |H_p^S| = 1; |X_p^S(h)| = 1; |G_p^S(x, h)| = 1; |Y_p^S(x, g, h)| = 1; \\ |Q_p^S(x, g, y, h)| = 1; |Z_p^S(x, g, y, q, h)| = 1; |R_p^S(x, g, y, q, z, h)| = 1, \end{aligned}$$

а логіко-лінгвістична модель (1)–(4) набуває вигляду:

$$L^S = \bigwedge_{p \in P^S} L_p^S(x, g, y, q, z, r, h),$$

кількість яких визначається потужністю множини відношень:

$$v^S = \sum_{p \in P^S} |R_p^S(x, g, y, q, z, h)| = |P^S|.$$

Наприклад, для простого речення природної мови, ускладненого однорідними присудками «Цікава лекція заохочує та спонукає студентів до подальшого навчання» типова форма логіко-лінгвістичної моделі (1)–(4) буде мати такий вигляд:

$$\begin{aligned} L(S) = L^S = p_1(x_1, g_1, y_1, q_1, z_1, r_1, h_1) \& p_2(x_1, g_1, y_1, q_1, z_1, r_1, h_1) = \\ = p_1(x_1, g_1, y_1, 0, z_1, r_1, 0) \& p_2(x_1, g_1, y_1, 0, z_1, r_1, 0), \end{aligned}$$

$$L(S) = L^S = \text{заохочує (лекція, цікава, студентів, 0, навчання, подальшого, 0)} \& \text{спонукає (лекція, цікава, студентів, 0, навчання, подальшого, 0)}.$$

Б) Якщо множина $C_p^S(x, h)$ характеристик (параметрів) хоча б одного із суб'єктів x містить більше одного елемента, то

$$\begin{aligned} |P^S| &= 1; |H_p^S| = 1; |Y_p^S(x, g, h)| = 1; |Q_p^S(x, g, y, h)| = 1; \\ |Z_p^S(x, g, y, q, h)| &= 1; |R_p^S(x, g, y, q, z, h)| = 1, \end{aligned}$$

а логіко-лінгвістична модель (1)–(4) буде мати такий узагальнений вигляд:

$$L^S = \bigwedge_{x \in X_p^S(h)} \bigwedge_{g \in G_p^S(x, h)} L_p^S(x, g, y, q, z, r, h),$$

кількість яких визначається:

$$v^S = \sum_{x \in X_p^S(h)} \sum_{g \in G_p^S(x, h)} |R_p^S(x, g, y, q, z, h)| = \sum_{x \in X_p^S(h)} |G_p^S(x, h)|.$$

Наприклад, для простого речення природної мови, ускладненого однорідними означеннями для підмета «Цікава та інформативна лекція заохочує студентів до подальшого навчання» типова форма логіко-лінгвістичної моделі (1)–(4) буде мати такий вигляд:

$$\begin{aligned} L(S) = L^S &= p_1(x_1, g_1, y_1, q_1, z_1, r_1, h_1) \& p_1(x_1, g_2, y_1, q_1, z_1, r_1, h_1) = \\ &= p_1(x_1, g_1, y_1, 0, z_1, r_1, 0) \& p_1(x_1, g_2, y_1, 0, z_1, r_1, 0), \end{aligned}$$

$L(S) = L^S$ = заохочує (лекція, цікава, студентів, 0, навчання, подальшого, 0) & заохочує (лекція, інформативна, студентів, 0, навчання, подальшого, 0).

В) Якщо множина $Q_p^S(x, g, y, h)$ характеристик (параметрів) хоча б одного із об'єктів у містить більше одного елемента, то $|P^S| = 1; |H_p^S| = 1; |X_p^S(h)| = 1; |G_p^S(x, h)| = 1; |Z_p^S(x, g, y, q, h)| = 1; |R_p^S(x, g, y, q, z, h)| = 1$, а логіко-лінгвістична модель (1) – (4) буде мати такий узагальнений вигляд:

$$L^S = L_p^S(h),$$

$$L_p^S(h) = L_p^S(x, g, h),$$

$$L_p^S(x, g, h) = \bigwedge_{y \in Y_p^S(x, g, h)} \bigwedge_{q \in Q_p^S(x, g, y, h)} L_p^S(x, g, y, q, h),$$

$$L_p^S(x, g, y, q, h) = L_p^S(x, g, y, q, z, r, h).$$

Логіко-лінгвістична модель (1)–(4) зводиться до кон'юнкції простих предикатів $L_p^S(x, g, y, q, z, r, h)$:

$$L^S = \bigwedge_{y \in Y_p^S(x, g, h)} \bigwedge_{q \in Q_p^S(x, g, y, h)} L_p^S(x, g, y, q, z, r, h),$$

кількість яких визначається:

$$\begin{aligned} v^S &= \sum_{y \in Y_p^S(x, g, h)} \sum_{q \in Q_p^S(x, g, y, h)} |R_p^S(x, g, y, q, z, h)| = \\ &= \sum_{y \in Y_p^S(x, g, h)} |Q_p^S(x, g, y, h)|. \end{aligned}$$

Наприклад, для простого речення природної мови, ускладненого однорідними означеннями та додатками «Цікава лекція заохочує лінивих і працюючих студентів та практикантів до подальшого навчання» типова форма логіко-лінгвістичної моделі (1)–(4) буде мати такий вигляд:

$$\begin{aligned} L(S) = L^S &= p_1(x_1, g_1, y_1, q_1, z_1, r_1, h_1) \& p_1(x_1, g_1, y_2, q_1, z_1, r_1, h_1) \& \\ &p_1(x_1, g_1, y_1, q_2, z_1, r_1, h_1) \& p_1(x_1, g_1, y_2, q_2, z_1, r_1, h_1) = \\ &= p_1(x_1, g_1, y_1, q_1, z_1, r_1, 0) \& p_1(x_1, g_1, y_2, q_1, z_1, r_1, 0) \& \\ &p_1(x_1, g_1, y_1, q_2, z_1, r_1, 0) \& p_1(x_1, g_1, y_2, q_2, z_1, r_1, 0), \end{aligned}$$

$L(S) = L^S$ = заохочує (лекція, цікава, студентів, лінивих, навчання, подальшого, 0) & заохочує (лекція, цікава, практикантів, лінивих, навчання, подальшого, 0) & заохочує (лекція, цікава, студентів, працюючих, навчання, подальшого, 0) & заохочує (лекція, цікава, практикантів, працюючих, навчання, подальшого, 0).

Г) Якщо множина $R_p^S(x, g, y, q, z, h)$ характеристик (параметрів) хоча б одного із предметів р-го відношення містить більше одного елемента, то

$$\begin{aligned} |P^S| &= 1; |H_p^S| = 1; |X_p^S(h)| = 1; |G_p^S(x, h)| = 1; \\ |Y_p^S(x, g, h)| &= 1; |Q_p^S(x, g, y, h)| = 1, \end{aligned}$$

а логіко-лінгвістична модель (1)–(4) буде мати такий узагальнений вигляд:

$$L^S = L_p^S(h),$$

$$L_p^S(h) = L_p^S(x, g, h),$$

$$L_p^S(x, g, h) = L_p^S(x, g, y, q, h),$$

$$L_p^S(x, g, y, q, h) = \bigwedge_{z \in Z_p^S(x, g, y, q, h)} \bigwedge_{r \in R_p^S(x, g, y, q, z, h)} L_p^S(x, g, y, q, z, r, h).$$

Логіко-лінгвістична модель (1)–(4) зводиться до кон'юнкції простих предикатів $L_p^S(x, g, y, q, z, r, h)$:

$$L^S = \bigwedge_{z \in Z_p^S(x, g, y, q, h)} \bigwedge_{r \in R_p^S(x, g, y, q, z, h)} L_p^S(x, g, y, q, z, r, h),$$

кількість яких визначається за формулою:

$$v^S = \sum_{z \in Z_p^S(x, g, y, q, h)} |R_p^S(x, g, y, q, z, h)|.$$

Наприклад, для простого речення природної мови, ускладненого однорідними додатками «Цікава лекція заохочує студентів до подальшого навчання та роботи» типова форма логіко-лінгвістичної моделі (1)–(4) буде мати такий вигляд:

$$\begin{aligned} L(S) = L^S &= p_1(x_1, g_1, y_1, q_1, z_1, r_1, h_1) \& p_1(x_1, g_1, y_1, q_1, z_2, r_1, h_1) = \\ &= p_1(x_1, g_1, y_1, 0, z_1, r_1, 0) \& p_1(x_1, g_1, y_1, 0, z_2, r_1, 0), \end{aligned}$$

$L(S) = L^S$ = заохочує (лекція, цікава, студентів, 0, навчання, подальшого, 0) & заохочує (лекція, цікава, студентів, 0, роботи, 0, 0).

6. Обговорення результатів дослідження типових форм логіко-лінгвістичних моделей

Перехід від узагальненої форми логіко-лінгвістичної моделі до типових форм дозволив виявити залежність кількості елементів множин складових логі-

ко-лінгвістичної моделі від типу ускладнення простих речень природної мови. Перехід від загальної формули (1) до предикату $L_p^s(x, g, y, q, z, r, h)$ є декомпозицією проблеми формального опису довільного речення природної мови та відображає системний підхід до її вирішення [2]. На відміну від описаної в [15] уніфікованої форми, логіко-лінгвістична модель (1)–(4) враховує можливі комбінації узгодження відношень, суб'єктів, об'єктів, предметів відношень та їх характеристик.

Таким чином, із виведених окремих форм видно, що формування логіко-лінгвістичних моделей для різних типів речень природної мови відбувається за єдиним принципом: уніфікована форма, шаблон, заповнюється для кожного речення предикатними змінними та константами у відповідності з тим, яку синтаксичну роль виконує те чи інше слово. Тобто зміст кожного речення природної мови фактично представляє собою визначення відношень між підметом (суб'єктом) та присудком (предикатом), а також їх спільне відношення до того, що вони виражають в дійсності, та формують основну граматичну властивість речення – його суть – предикативність.

7. Висновки

Аналіз речень природної мови дав змогу виявити концептуальні відношення узгодження, керування та

прилягання, на основі чого сформовано логіко-лінгвістичну модель з функціональними зв'язками. Для виявлення концептуальних відношень між словами в реченні природної мови використовується система продукцій, що представляє собою набір правил типу «modus ponens».

Виявлено можливі типи зв'язків у реченнях природної мови, що відображаються у логіко-лінгвістичній моделі за допомогою логічних операцій кон'юнкції, диз'юнкції та імплікації.

Залежно від концептуальних зв'язків між простими реченнями, які входять до складного, частини логіко-лінгвістичної моделі (1)–(4) інтерпретуються, як складні або прості висловлювання, що дозволяє рекурсивно повертатися до загального вигляду формули (1)–(4) і накладати її на конкретну ситуацію доти, доки не будуть однозначно інтерпретовані всі частини речення природної мови S , що відображають закінчений зміст.

Створення узагальненої логічної моделі представлення знань та її типових форм з використанням апарату логіки предикатів дозволило розробити єдиний підхід до змістовного аналізу речень природної мови, що продемонстровано на прикладах. Це стало можливим завдяки виявленню можливих типів логічних зв'язків у синтаксичних конструкціях речень природної мови.

Література

1. Башмаков, А. И. Интеллектуальные информационные технологии [Текст]: учеб. пос. / А. И. Башмаков, И. А. Башмаков. – М.: МГТУ им. Баумана, 2005. – 304 с.
2. Вавіленкова, А. І. Теоретичні основи аналізу електронних текстів [Текст]: монографія / А. І. Вавіленкова, Д. В. Ланде, О. Є. Литвиненко. – К.: НАУ, 2015. – 258 с.
3. Gries, S. Th. Corpus-based methods and cognitive semantics: the many meanings of to run [Text] / S. Th. Gries. – Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis, 2006. – P. 57–99.
4. Evans, V. Lexical concepts, cognitive models and meaning-construction [Text] / V. Evans // Journal of Cognitive semiotics. – 2006. – P. 73–107.
5. Кобозева, И. М. Лингвистическая семантика [Текст] / И. М. Кобозева. – М.: Эдитореал УРСС, 2000. – 352 с.
6. Поспелов, Д. А. Логико-лингвистические модели в системах управления [Текст] / Д. А. Поспелов. – М.: Энергоиздат, 1981. – 232 с.
7. Осуга, С. Приобретение знаний [Текст] / С. Осуга, Ю. Саэки. – М.: Мир, 1990. – 304 с.
8. Джарратано, Д. Экспертные системы: принципы разработки и программирование; 4-е изд. [Текст]: пер. с англ. – М.: ООО «Вильямс», 2007. – 1152 с.
9. Широков, В. А. Корпусна лінгвістика [Текст] / В. А. Широков, О. В. Бугаков, Т. О. Грязнухіна та ін. – К.: Довіра, 2005. – 471 с.
10. Kazuhisa, S. Design of a conceptual level programming environment based on task ontology [Text] / S. Kazuhisa, I. Mitsuru, K. Osamu, M. Riichiro // Proc. of Successes and failures of knowledge based systems in real world applications, 1996. – P. 11–22.
11. Jiawei, H. Data Mining: Concepts and Techniques [Text] / H. Jiawei, K. Micheline. – Morgan Kaufmann Publishers, 2000.
12. Cross, D. Data Munging with Perl [Text] / D. Cross // Manning Publication Co, USA, 2001.
13. Jurafsky, D. Natural Language Processing [Електронний ресурс] / D. Jurafsky, C. Manning. – Stanford University, 2012. – Available at: <https://www.coursera.org/course/nlp/> – 15.05.2014
14. Вавіленкова, А. І. Методологічні основи автоматичного аналізу логіко-лінгвістичних моделей текстових документів [Текст] / А. І. Вавіленкова // Математичні машини та системи. – 2015. – № 1. – С. 65–71.
15. Вавіленкова, А. И. Извлечение смысла из предложений естественного языка [Текст] / А. И. Вавіленкова // Программные продукты и системы. – 2012. – № 4 (100). – С. 87–90.